



This is a repository copy of *Detecting and tracking time-varying causality with applications to EEG data*.

White Rose Research Online URL for this paper:  
<http://eprints.whiterose.ac.uk/74675/>

---

### **Monograph:**

Zhao, Y., Billings, S.A., Wei, H.L. et al. (1 more author) (2011) Detecting and tracking time-varying causality with applications to EEG data. Research Report. ACSE Research Report no. 1027 . Automatic Control and Systems Engineering, University of Sheffield

---

### **Reuse**

Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

### **Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>

# Detecting and Tracking Time-varying Causality with Applications to EEG Data

Y. Zhao, S.A. Billings, H. L. Wei, P.G. Sarrigiannis



Research Report No. 1027

Department of Automatic Control and Systems Engineering  
The University of Sheffield  
Mappin Street, Sheffield,  
S1 3JD, UK

October, 2011

# Detecting and Tracking Time-varying Causality with Applications to EEG Data

Y. Zhao, S. A. Billings, H. L. Wei, P.G. Sarrigiannis

**Abstract**—This paper introduces a novel method called the ERR-Causality, or Error Reduction Ratio Causality test, that can be used to detect and track causal relationships between two signals using a new adaptive forward orthogonal least squares (Adaptive-Forward-OLS) algorithm. In comparison to the traditional Granger method, one advantage of the new ERR-Causality test is that it can effectively detect the time-varying direction of linear or nonlinear causality between two signals without fitting a complete model. Another important advantage is that the ERR-Causality test can detect both the direction of interactions and estimate the relative time shift between the two signals. Several numerical examples are provided to illustrate the effectiveness of the new method for causal relationship detection between two signals. An important real application, relating to the analysis of the causality of EEG signals from different cortical sites which can be very useful for understanding brain activity during an epileptic seizure by inspecting the high-resolution time-varying directed information flow, is also discussed.

**Index Terms**—Causality, Granger, EEG, Time-varying, OLS

## I. INTRODUCTION

The detection of hidden interdependencies between the components of complex dynamic systems is an important problem that arises in many research fields. There are several ways to tackle this problem based on using either an explicit generative model that embraces the known nonlinear causal architecture [1], or by simply establishing statistical dependencies between two signals using coherence, phase synchronization, or the Granger causality test. The latter approaches are usually more viable and many methods based on these ideas have been developed recently and applied to the analysis of electrophysiological signals, such as directed coherence and partial directed coherence [2, 3, 4]. However cross correlation methods have two possible drawbacks, one is the requirement for reasonably long data sets, and the other is that correlation may only detect causality of two signals with linear interactions. The minimal required data window size to achieve correct results is important because too wide a window will decrease the temporal resolution of the analysis, which can be fatal if the casual relationship changes rapidly over time, and too narrow

a window reduces the statistical reliability.

Signals sampled from the real world are rarely stationary and well behaved, and casual interactions and couplings can typically appear, disappear and reappear, and may become weaker or grow stronger over time. Moreover, most complex systems exhibit nonlinear dynamic behaviours, which may lead to a possible failure of the cross correlation method. Another established way to solve the causality detection problem is by mutually predicting selected observable measurements based on multivariate autoregressive modelling. Many methods based on this idea have been proposed recently and one of the best established methods is based on the Granger causality test [5]. The key idea of this method is that if a signal X causes a signal Y, the knowledge of the past of both X and Y should improve the prediction of the presence of Y in comparison with the knowledge of the past of Y alone. Many new methods have been developed which extend this idea [6, 7, 8]. However, all these methods require that the system model is fully known or that an unbiased model can be fitted to the data sets before the Granger test can be applied. This is far from straightforward when the underlying system relationships are nonlinear and dynamic and the measured observations are noisy because, unless a complete and full model which accounts for any potentially nonlinear noise effects is estimated, the Granger test results will be compromised.

In the present study a new causality test is introduced which overcomes most of the disadvantages of existing methods. The new causality detection method will be referred to as the ERR-Causality or Error Reduction Ratio-Causality test. The key advantage of the new test is that it can be applied to nonlinear dynamic systems, and unlike the Granger based tests, the new method does not depend on the full knowledge or estimation of a complete and unbiased system model. By exploiting an important property of the error reduction ratio (ERR) test that is part of the orthogonal least squares (OLS) algorithm it is shown that the causal flow can be detected even when the model is incomplete. This is a significant advantage when the underlying system is nonlinear and dynamic and the measurements may be noisy, because a complete and full model including a nonlinear noise model, which would normally be required to yield unbiased model estimates is not required and indeed not even the full parameter estimates are used in the test. These advantages mean the test is relatively easy to apply and can be used to track fast transitions between causal effects to detect the direction of linear or nonlinear casual interactions, the strength of these interactions, and to provide an estimate of the time shift between two directional signals. Three numerical examples are

Department of Automatic Control and System Engineering, University of Sheffield, UK.

Department of Clinical Neurophysiology, Sheffield Teaching Hospitals NHS Foundation Trust, Royal Hallamshire Hospital

used to illustrate the application of the new test and to show the performance of the method in comparison with other methods. Finally the application of the new method to real high resolution Electroencephalography (EEG) recordings is described and it is shown how the new test can be used to exploit the flow sequence of brain signals which may help to locate the source [9] and understand brain activity during an epileptic seizure.

## II. METHODS

Let  $X = \{x(t)\}$  and  $Y = \{y(t)\}$  be two signals,  $t = 1, \dots, M$ , where  $M$  is the data length. The aim of this paper is to measure the casual interaction over time between these two signals. The results can be, for example, at a specific time, the signal  $X$  causes  $Y$ ,  $Y$  causes  $X$ , no interaction or bi-directional interaction between them. For a complex system, the causality is often time-varying and the interaction is often dynamic and nonlinear, which makes the problem more challenging. This section begins with a brief review of the cross correlation and the Granger causality tests associated with this problem, and then presents a new ERR-Causality test.

### A. Cross correlation

The cross correlation is the most commonly used method to detect causal interactions between the signals  $X$  and  $Y$ , and is defined as

$$\phi_{xy}(\tau) = \frac{\sum_{t=\tau+1}^M (x(t) - \bar{x})(y(t + \tau) - \bar{y})}{\left[ \sum_{t=1}^M (x(t) - \bar{x})^2 \sum_{t=1}^M (y(t) - \bar{y})^2 \right]^{1/2}}, \quad (1)$$

where  $\tau = 0, \pm 1, \pm 2, \dots, \pm(M-1)$  and  $\bar{x}, \bar{y}$  denote the means of  $X, Y$  respectively. If a well defined peak at lag  $\tau$  can be observed in the cross correlation function, this indicates that the signal  $X$  lags behind the signal  $Y$  if  $\tau > 0$ , which means that  $X$  causes  $Y$  from the causality point of view. Or if the signal  $X$  lags the signal  $Y$  when  $\tau < 0$ , this means  $Y$  causes  $X$ . If no well defined peak can be observed, no causality is detected. This method is easy to understand and to implement, and no knowledge of the exact model underlying the interaction is required. However, there are three potential problems. Firstly, correlation may not detect the nonlinear causality between two signals, however most complex real systems are likely to be nonlinear. Secondly, this method can not detect a complicated causality, such as a bi-directional interaction. Thirdly, this method requires relatively long data sets to achieve accurate results, which means the reaction to rapidly changing causality over time is relatively slow.

### B. Granger Causality

A well established approach to detect causality in both linear and nonlinear systems is the Granger causality test [5]. To calculate the Granger causality of  $X$  to  $Y$ , a model has to be pre-established which defines the relationship between the output  $Y$  and its past information  $Y^-$  and the past information of the input  $X^-$ , expressed as:

$$Y = f(Y^-, X^-) \quad (2)$$

Based on the sampled data the parameters in the model  $f(Y^-, X^-)$  have to be estimated and then the predictions of  $Y$  based on  $Y^-$  alone, and on  $Y^-$  and  $X^-$  are generated. In both cases, the accuracy of prediction may be expressed by the variance of the prediction errors for two-dimensional modelling  $\text{var}(Y|Y^-)$ ,  $\text{var}(Y|Y^-, X^-)$ . The Granger causality of  $X$  to  $Y$ ,  $G_{X \rightarrow Y}$ , is defined by

$$G_{X \rightarrow Y} = \ln \frac{\text{var}(Y|Y^-)}{\text{var}(Y|Y^-, X^-)} \quad (3)$$

The Granger causality of  $Y$  to  $X$  is defined by

$$G_{Y \rightarrow X} = \ln \frac{\text{var}(X|X^-)}{\text{var}(X|X^-, Y^-)} \quad (4)$$

The advantages of this method are that if the model structure is chosen appropriately, it can tackle both linear and nonlinear systems. The test is also able to detect a bi-directional causality because the causalities from  $Y$  to  $X$  and from  $X$  to  $Y$  are calculated separately. The required window size of sampled data depends on the dynamical properties of the original signals, and the complexity of the chosen model structure. One possible problem for this method is that if the model structure is not chosen appropriately, for example, the model is missing a significant term or terms, the calculated Granger causality may not be reliable, which will be demonstrated in the second simulation example in section III-B. Because this test is based on an accurate model then noise models may need to be estimated to ensure the model is unbiased. For nonlinear relationships this will often require a nonlinear noise model. Fitting complete nonlinear dynamics system and noise models is a significant overload for this test. Moreover, this method can only detect the direction of signal flow, but is not able to provide a quantitative insight into the time shift between the two signals.

### C. Adaptive-Forward-OLS

The orthogonal least squares (OLS) algorithm [10, 11, 12], is a popular approach that has been widely used in nonlinear system identification where the orthogonal least squares searches through all the possible candidate

model terms to select the most significant model terms which are then included to build models term by term. The significance of each of the selected model terms is measured by an index, called the error reduction ratio (ERR), which indicates how much (in percentages) of the variance change in the system response can be accounted for by including the relevant model terms. Complex nonlinear dynamic models and nonlinear noise models can all be identified using this algorithm.

This section introduces an Adaptive-Forward-OLS algorithm which will be used later by modifying the well known forward-regression version of OLS [10]. Consider the linear regression function

$$y(t) = \sum_{i=1}^N p_i(t)\theta_i, t = 1, \dots, M \quad (5)$$

where  $y(t)$  is the dependent variable or the term to regress upon,  $p_i(t)$  are regressors,  $\theta_i$  are unknown parameters to be estimated and  $M$  denotes the number of data points in the data set. Equation (5) can be written as

$$Y = P\Theta \quad (6)$$

where

$$Y = \begin{bmatrix} y(1) \\ \vdots \\ y(M) \end{bmatrix}, P = \begin{bmatrix} P^T(1) \\ \vdots \\ P^T(M) \end{bmatrix}, \Theta = \begin{bmatrix} \theta(1) \\ \vdots \\ \theta(N) \end{bmatrix} \quad (7)$$

and

$$P^T(t) = (p_1(t), \dots, p_N(t)) \quad (8)$$

Matrix  $P$  can be decomposed as  $P = W \times A$  where

$$W = \begin{bmatrix} w_1(1) & \dots & w_N(1) \\ \vdots & & \vdots \\ w_1(M) & \dots & w_N(M) \end{bmatrix} \quad (9)$$

is an orthogonal matrix because

$$W^T W = \text{Diag} \left[ \sum_{t=1}^M w_1^2(t), \dots, \sum_{t=1}^M w_N^2(t) \right] \quad (10)$$

and  $A$  is an upper triangular matrix with unity diagonal elements

$$A = \begin{bmatrix} 1 & a_{12} & a_{13} & \cdots & a_{1N} \\ & 1 & a_{23} & \cdots & a_{2N} \\ & & \ddots & \ddots & \vdots \\ & & & 1 & a_{N-1N} \\ & & & & 1 \end{bmatrix} \quad (11)$$

Therefore, (6) can be rewritten as

$$Y = WG \quad (12)$$

where

$$G = A\Theta = [g_1, \dots, g_N]^T \quad (13)$$

The estimation of the original parameters can be computed from

$$\left. \begin{aligned} \hat{\theta}_N &= \hat{g}_N \\ \hat{\theta}_i &= \hat{g}_i - \sum_{k=i+1}^N a_{ik}\hat{\theta}_k, i = N-1, \dots, 1 \end{aligned} \right\} \quad (14)$$

In traditional forward OLS, the cut off value of ERR,  $C_{off}$ , to stop the search procedure and determine the number of significant terms can be difficult to select, especially when the level of noise is unknown. Recently, several criteria based on ERR have been developed to monitor and stop the search procedure [13, 14, 15]. This paper introduces an algorithm named the Adaptive-Forward-OLS by utilizing the penalized error-to-signal ratio

$$PESR_n = \frac{1}{(1 - \lambda n/M)^2} \left( 1 - \sum_{i=1}^n [err]_i \right) \quad (15)$$

to monitor the regressor search procedure, where  $n$  denotes the number of selected terms and  $M$  denotes the total number of sampled data. The search procedure stops when  $PESR_n$  arrives at a minimum. The effect of the adjustable parameter  $\lambda$  on the results is discussed in [16], which suggested that  $\lambda$  should be chosen between 5 and 10. The value of  $\lambda$  is chosen as 6 for all the examples in this paper based on experience, but other values in this range have also been tested and the results remained correct and unchanged.

The whole procedure of the Adaptive-Forward-OLS algorithm can be summarized as follows.

- (a)  $a_{11} = 1$ ,  $w_1(t) = p_1(t)$ , and  $\hat{g}_1 = \frac{\sum_{t=1}^M w_1(t)y(t)}{\sum_{t=1}^M w_1^2(t)}$ .
- (b) For  $k = 2, \dots, N$ :  $a_{ik} = \frac{\sum_{t=1}^M w_i(t)p_k(t)}{\sum_{t=1}^M w_i^2(t)}$ ,  $i = 1, \dots, k-1$ ,  $a_{kk} = 1$ ,  $w_k(t) = p_k(t) - \sum_{i=1}^{k-1} a_{ik}w_i(t)$ , and  $\hat{g}_k = \frac{\sum_{t=1}^M w_k(t)y(t)}{\sum_{t=1}^M w_k^2(t)}$ . ERR is used as a criterion for model structure selection, and is defined as

$$[err]_k = \frac{\hat{g}_k^2 \sum_{t=1}^M w_k^2(t)}{\sum_{t=1}^M y^2(t)} \quad (16)$$

- (c) Compute  $PESR_k$  using (15). The search procedure stops when  $PESR_k$  arrives at a minimum.

Noise modelling which will often be required to ensure unbiased models, and is described in [17, 18].

#### D. A New Granger Causality Test based on the Adaptive-Forward-OLS algorithm

In this section a new modification of the Granger test will be introduced based on the modelling algorithm in

section II-C above. When applying the Granger test, the model is either pre-known, which is often impossible especially for real systems, or the model structure has to be detected and a model estimated as the initial step. It has been shown [19] that the estimator introduced in section II-C combines structure determination, parameter estimation and noise modelling, and when coupled with model validity tests, is particularly powerful in identifying parsimonious models for structure-unknown systems. The Adaptive-Forward-OLS algorithm therefore can be a part of the Granger method to improve the identification performance and hence enhance the detection capability of the Granger test.

#### E. The ERR-Causality Test

The new ERR-Causality test is introduced in this section by tackling the problem another way to detect causality between two signals without the identification of a full or complete model. It is shown that this test has significant advantages compared to existing tests and can be readily applied to linear and nonlinear dynamic systems even with noise corrupted measurements.

Consider a bivariate Autoregressive (ARX) model

$$\begin{aligned} x(t) &= \sum_{i=1}^{p_x} a_i x(t-i) + \sum_{j=1}^{p_y} c_j y(t-j) + e_x(t) \\ y(t) &= \sum_{i=1}^{q_y} b_i y(t-i) + \sum_{j=1}^{q_x} d_j x(t-j) + e_y(t) \end{aligned} \quad (17)$$

where  $e_x(t), e_y(t)$  denote noise sequences, which can be either white noise or coloured noise. Obviously, if  $c_j \neq 0, j \in \{1, \dots, p_y\}$  and  $d_j \neq 0, j \in \{1, \dots, q_x\}$ , this is a typical bi-directional system, which means  $X$  causes  $Y$ , and at the same time  $Y$  causes  $X$ . Consider initially the causality from  $X$  to  $Y$ . A NARX (Non-linear Auto-Regressive with eXogenous inputs) model [20] constructed using basic function expansions using a linear-in-the parameters form is introduced to express  $Y$

$$y(t) = \sum_{i=1}^N \theta_i \phi_i(t) + e(t) \quad (18)$$

where  $\theta_i$  are unknown parameters,  $N$  is the number of the total potential model terms involved, and  $\phi_i(t) = \phi_i(\varphi(t))$  are model terms generated from a candidate term set, for example,  $\varphi(t)$  can be

$$\varphi(t) = \{1, y(t-1), \dots, y(t-n_y), x(t-1), \dots, x(t-n_x)\}^T \quad (19)$$

which includes some simple linear components from the past information of  $X$  and  $Y$ .

Instead of generating a complete model that has to pass the validity tests, the ERR-Causality test can be summarized in the following.

*Initially, construct a candidate term set which typically includes past information of  $Y$ , and past information*

*of  $X$ . Apply the Adaptive-Forward-OLS algorithm and computer ERR and PESR values. If the selected significant terms by the Adaptive-Forward-OLS algorithm in section II-C includes any term from the past information of  $X$ , this indicates the signal  $X$  causes  $Y$  during the considered time duration  $[t - h/2, t + h/2]$ , where  $h$  denotes the sampling window size. The ERR-Causality from  $X$  to  $Y$  at time  $t$ , expressed as  $F_{X \rightarrow Y}(t)$ , is then defined as 1. If no component from the past information of  $X$  is included in the selected significant terms, this indicates that  $X$  has no interaction with  $Y$  during  $[t - h/2, t + h/2]$ , and  $F_{X \rightarrow Y}(t)$  is defined to be 0. The strength of  $F_{X \rightarrow Y}(t)$  can be estimated by the summed ERR values of all the selected terms from  $X^-$ , the maximum strength being 1.*

The selection of the candidate term set can be much more complicated than (19), and will depend on the pre-known information of the considered system. For example, (20) shows a candidate term set with some non-linear components.

$$\varphi(t) = \left\{ \begin{aligned} &y(t-1), y(t-2), \dots, y(t-n_y), \\ &x(t-1), x(t-2), \dots, x(t-n_x), \\ &y(t-1)x(t-1), \dots, y(t-1)x(t-n_x), \\ &y^2(t-1), y^2(t-2), \dots, y^2(t-n_y), \\ &x^2(t-1), x^2(t-2), \dots, x^2(t-n_x) \end{aligned} \right\}^T \quad (20)$$

If any significant term that includes any component from the past information of  $X$  is chosen in the ERR-Causality test, this method, theoretically, is able to observe the causality from  $X$  to  $Y$ , even though  $\varphi(t)$  may not include a complete set of all the correct terms of the system. This advantage is based on the fact that the order of the ERR values or the order of term selection produced by the Adaptive-Forward-OLS algorithm is correct even when a complete model is not estimated. This important result, which is fundamental to the ERR-Causality test, will be proved next.

Consider the model

$$y(t) = \sum_{i=1}^N w_i(t) g_i + \zeta(t), t = 1, \dots, M \quad (21)$$

where the first  $N$  terms represent all the correct model terms and  $\zeta(t)$  is a white noise sequence with zero mean. Assume only  $N_p$  terms are selected and the other terms are not considered in  $\varphi(t)$ . Note that noise terms can be included in the  $N$  terms in the model (21), which should then reduce  $\zeta(t)$  to be white. Then (21) can be expressed as

$$y(t) = \sum_{i=1}^{N_p} w_i(t) g_i + \sum_{j=N_p+1}^N w_j(t) g_j + \zeta(t) \quad (22)$$

Now (22) can be rewritten as

$$y(t) = \sum_{i=1}^{N_p} w_i(t)g_i + e(t) \quad (23)$$

where

$$e(t) = \sum_{j=N_p+1}^N w_j(t)g_j + \zeta(t) \quad (24)$$

represents missing model terms and can be viewed as coloured noise and may not be zero mean. Squaring both sides of (23) and taking the expected value gives

$$\begin{aligned} E[y^2(t)] &= E\left[\sum_{i=1}^{N_p} w_i^2(t)g_i^2\right] + \\ &2E\left[\sum_{i=1}^{N_p} w_i(t)g_i e(t)\right] + \\ &E[e^2(t)] \end{aligned} \quad (25)$$

Obviously,

$$E\left[\left(\sum_{i=1}^{N_p} w_i(t)g_i\right)^2\right] = E\left[\sum_{i=1}^{N_p} w_i^2(t)g_i^2\right] \quad (26)$$

because  $w(i)$  are orthogonal,  $w(i)w(j) = 0$  ( $i \neq j$ ).

Then (25) can be rewritten as

$$\frac{1}{M} \sum_{t=1}^M y^2(t) - \sum_{i=1}^{N_p} \frac{1}{M} \sum_{t=1}^M w_i^2(t)g_i^2 = \alpha \quad (27)$$

where

$$\alpha = 2E\left[\sum_{i=1}^{N_p} w_i(t)g_i e(t)\right] + E[e^2(t)] \quad (28)$$

Replacing  $e(t)$  in (28) by (24)

$$\begin{aligned} \alpha &= 2E\left[\sum_{i=1}^{N_p} w_i(t)g_i \left(\sum_{j=N_p+1}^N w_j(t)g_j + \zeta(t)\right)\right] \\ &+ E\left[\left(\sum_{j=N_p+1}^N w_j(t)g_j + \zeta(t)\right)^2\right] \\ &= 0 + 0 + E\left[\sum_{j=N_p+1}^N w_j^2(t)g_j^2\right] + E[\zeta^2(t)] \\ &= \frac{1}{M} \sum_{t=1}^M \sum_{j=N_p+1}^N w_j^2(t)g_j^2 + \frac{1}{M} \sum_{t=1}^M \zeta^2(t) \end{aligned} \quad (29)$$

Substituting  $\alpha$  back into (27) and dividing  $\frac{1}{M} \sum_{t=1}^M y^2(t)$  to both side produces

$$\begin{aligned} 1 - \frac{\sum_{i=1}^{N_p} \frac{1}{M} \sum_{t=1}^M w_i^2(t)g_i^2}{\frac{1}{M} \sum_{t=1}^M y^2(t)} &= \\ \frac{\frac{1}{M} \sum_{t=1}^M \sum_{j=N_p+1}^N w_j^2(t)g_j^2 + \frac{1}{M} \sum_{t=1}^M \zeta^2(t)}{\frac{1}{M} \sum_{t=1}^M y^2(t)} \end{aligned} \quad (30)$$

Based on the definition of ERR in (16), finally this yields

$$1 - \sum_{i=1}^{N_p} [err]_i = \sum_{j=N_p+1}^N [err]_j + \frac{\sigma_\zeta^2}{\sigma_y^2} \quad (31)$$

where  $\sigma_\zeta, \sigma_y$  denote the standard deviation of  $\zeta(t), y(t)$  respectively. Equation (31) implies that the ERR values for the selected terms can be calculated quite independently of the un-selected terms of the correct term set. Hence, the proposed ERR-Causality method can always provide a correct order of significant terms without fitting a complete model. In other words, even when not all the significant terms are included in  $\varphi(t)$ , this method should still detect the causality. Notice also that it is only the ERR values that are used in the test, there is no requirement to fit a complete model or no requirement to even estimate all the model parameters. The ERR-Causality test is therefore a more powerful and robust causality detection method, and moreover, many time consuming calculations can be considerably avoided or reduced because the search procedure is monitored by PESR and no further parameter estimation is required. Notice that the method automatically defaults to select just linear terms if the relationship is linear but can also accommodate complete nonlinear dynamics relationships without full model estimation.

### III. SIMULATION STUDIES

This section discusses the efficiency and performance of the proposed new method by comparing results with the cross correlation and the Granger test method. The first example demonstrates the procedure including how to select significant terms and measure the value of ERR-Causality. The second and third examples demonstrate the advantages of the new method and the flexibility of term selection and reaction speed to causality changing over time.

#### A. Example 1

Consider an ARX model expressed as

$$y(t) = b_1 y(t-1) + b_2 y(t-2) + d_1 x(t-1) + d_2 x(t-2) + \zeta(t) \quad (32)$$

where  $\zeta(t)$  is a white noise of zero mean and a standard deviation  $\sigma = 0.05$ . In the first test, the parameters were set as  $b_1 = -0.6, b_2 = 0.2, d_1 = 0.2, d_2 = 0.1$ , which indicates  $Y$  depends on the past information of itself and  $X$ , or from the causality point of view,  $Y$  is caused by  $X$ . The model was simulated by setting the signal  $x(t)$  as a random sequence uniformly distributed in  $[-0.5, 0.5]$  and 1000 data point were collected after the system behaviour had settled down. The initial candidate term set  $\varphi(t)$  was chosen as

$$\{1, y(t-1), y(t-2), y(t-3), x(t-1), x(t-2), x(t-3)\}^T \quad (33)$$

The results of term selection associated with the PESR values from the Adaptive-Forward-OLS algorithm are

TABLE I  
THE LIST OF SORTED TERMS ASSOCIATED WITH PESR VALUES FOR  
THE FIRST TEST OF EXAMPLE 1.

Order	Terms	$[err]_i$	PESR
1	$x(t-1)$	0.33937	0.66863
2	$y(t-1)$	0.28525	0.38456
3	$x(t-2)$	0.09888	0.28673
<b>4</b>	<b><math>y(t-2)</math></b>	<b>0.02875</b>	<b>0.26009</b>
5	1	0.00140	0.26182
6	$x(t-3)$	0.00069	0.26435
7	$y(t-3)$	0.00021	0.26743

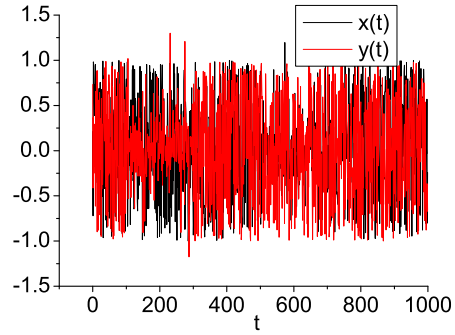
shown in TABLE I. It can be clearly seen that PESR arrives at the minimum 0.26009 when the number of selected terms is 4, which indicates the first 4 selected terms are significant and all the others can be discarded. Note in practice when using the Adaptive-Forward-OLS algorithm, the search procedure will stop after the first 5 terms, but the results of all terms are listed in TABLE I to show the trend of PESR in more detail. Because the past information of  $X$ ,  $x(t-1)$  and  $x(t-2)$  are included in the selected terms, the value of ERR-Causality from  $X$  to  $Y$ ,  $F_{X \rightarrow Y}$ , is detected as 1, and the corresponding strength is 0.43825 (the sum of ERR values for  $x(t-1)$  and  $x(t-2)$ ). Notice that the maximum strength is 1 because the ERR values for all terms sum to 1.

In the second test, the parameters were set as  $b_1 = -0.6$ ,  $b_2 = 0.2$ ,  $d_1 = 0$ ,  $d_2 = 0$ , which indicates there is no interaction between  $X$  and  $Y$ . The initial conditions and the candidate term set were exactly the same as those in the first test and the results are shown in TABLE II. It is shown that PESR arrives at the minimum 0.38921

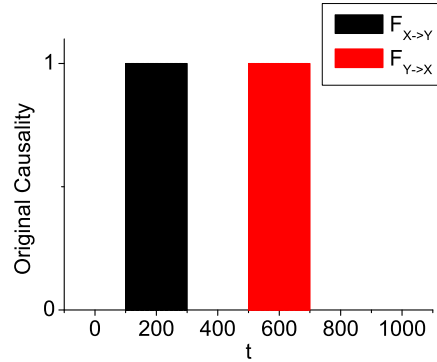
TABLE II  
THE LIST OF SORTED TERMS ASSOCIATED WITH PESR VALUES FOR  
THE SECOND TEST OF EXAMPLE 1

Order	Terms	$[err]_i$	PESR
1	$y(t-1)$	0.60121	0.40362
<b>2</b>	<b><math>y(t-2)</math></b>	<b>0.01886</b>	<b>0.38921</b>
3	$x(t-2)$	0.00067	0.39329
4	$x(t-1)$	0.00031	0.39781
5	$x(t-3)$	0.00010	0.40264
6	$y(t-3)$	0.00006	0.40761
7	1	0.00005	0.41268

when the number of selected terms is 2, which indicates only the first 2 terms are significant and all others can be discarded. Because the selected terms contain no past information of  $X$ , based on the definitions in this paper, the value of ERR-Causality from  $X$  to  $Y$ ,  $F_{X \rightarrow Y}$  is detected as 0 with a strength of 0.



(a)



(b)

Fig. 1. (a) The original signal  $x(t)$  and  $y(t)$  for Example 2. (b) The true causality of the signal  $X$  to  $Y$  and the signal  $Y$  to  $X$  over time, which shows  $F_{X \rightarrow Y} = 1$  during interval 100–300 and  $F_{Y \rightarrow X} = 1$  during interval 500–700.

## B. Example 2

This example aims to demonstrate the flexibility of the proposed new method in term selection in comparison with that of the Granger method. A total number of 1000 data points were generated using a time-varying model based on the definition in TABLE III, where  $\zeta_y(t)$ ,  $\zeta_x(t)$  were white noise sequences with zero mean and a standard deviation  $\sigma = 0.1$ , and  $r(t)$  denotes a random data sequence uniformly distributed in  $[-1, 1]$ . To save space, the notation  $y(t-1)$  is simplified as  $y^{-1}$ , and so on. Fig. 1.(a) shows the simulated signals of  $X$  and  $Y$ , and the true time-varying causality is shown in Fig. 1.(b), both of which and the model clearly indicate: the signal  $X$  causes  $Y$  at time 100–300; the signal  $Y$  causes  $X$  at time 500–700; with no causality at other times.

In the first test the Granger test was applied. Considering the causality from  $X$  to  $Y$ , this method used a nonlinear



TABLE III  
THE TIME-VARYING MODEL FOR EXAMPLE 2

$t$	$x(t)$	$y(t)$
0 – 100	$r(t)$	$r(t)$
101 – 300	$r(t)$	$-0.07x^{-1} + 0.32x^{-2} - x^{-1}x^{-2} + \zeta_y(t)$
301 – 500	$r(t)$	$r(t)$
501 – 700	$-0.07y^{-1} + 0.32y^{-2} - y^{-1}y^{-2} + \zeta_x(t)$	$r(t)$
701 – 1000	$r(t)$	$r(t)$

model

$$\begin{aligned}
 y(t) = & a(t) + \sum_{i=1}^3 b_i(t)y(t-i) + \sum_{i=1}^3 c_i(t)x(t-i) \\
 & + \sum_{i=1}^3 \sum_{j=i}^3 d_{ij}(t)y(t-i)y(t-j) \\
 & + \sum_{i=1}^3 \sum_{j=i}^3 f_{ij}(t)x(t-i)x(t-j) \\
 & + \sum_{i=1}^3 \sum_{j=i}^3 h_{ij}(t)x(t-i)y(t-j) + \zeta_y(t)
 \end{aligned} \quad (34)$$

where parameters  $a, b, c, d, f, h$  are no longer constants, but functions of time. The window size was chosen as 50 and the detected time-varying Granger causality is illustrated in Fig. 2.(a), where an accurate causality can be achieved if an appropriate threshold is used. This is essentially all implementation of the new algorithm described in section II-D.

Now a scenario when some key model terms are missed in the model structure is considered. By removing all the nonlinear terms in (34), the model

$$\begin{aligned}
 y(t) = & a(t) + \sum_{i=1}^3 b_i(t)y(t-i) + \sum_{i=1}^3 c_i(t)x(t-i) \\
 & + \zeta_y(t)
 \end{aligned} \quad (35)$$

was used and the detected time-varying Granger causality is illustrated by Fig. 2.(b), where the measurability of causality is significantly decreased and accurate results can not be achieved. This failure arises due to the large contribution of the term  $x(t-1)x(t-2)$  in the original model. The absence of this term in the model can dramatically change the signal to noise ratio (SNR) in  $Y$  and results in a very poor prediction of  $Y$  even when the past information of  $X$  is considered. This shows the disadvantage of any test based on a knowledge of a full and complete model.

The second test applied the ERR-Causality method.

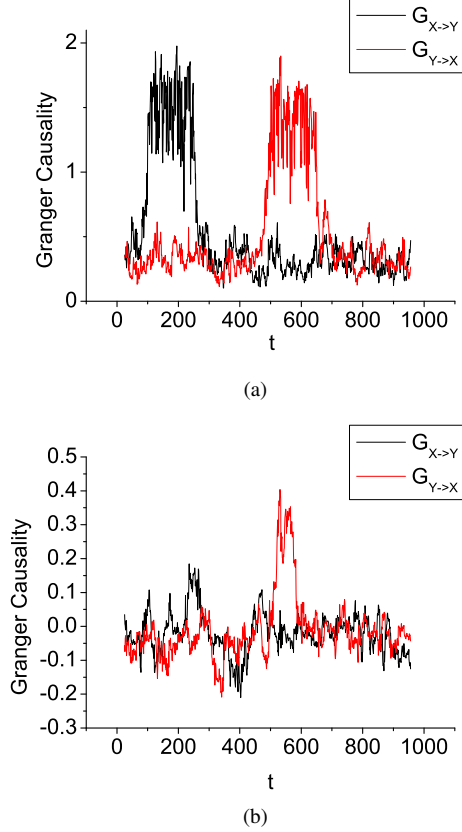


Fig. 2. (a) The detected time-varying Granger causality based on the model (34) for Example 2, where the causality is distinctive. (b) The detected time-varying Granger causality based on the model (35) for Example 2, where the causality is not distinctive.

Initially, the candidate terms set was chosen as

$$\begin{aligned}
 \{ & 1, \\
 & x(t-1), x(t-2), x(t-3), y(t-1), y(t-2), y(t-3), \\
 & x^2(t-1), x(t-1)x(t-2), x(t-1)x(t-3), x^2(t-2), \\
 & x(t-2)x(t-3), x^2(t-3), \\
 & y^2(t-1), y(t-1)y(t-2), y(t-1)y(t-3), y^2(t-2), \\
 & y(t-2)y(t-3), y^2(t-3), \\
 & x(t-1)y(t-1), x(t-1)y(t-2), x(t-1)y(t-3), \\
 & x(t-2)y(t-1), x(t-2)y(t-2), x(t-2)y(t-3), \\
 & x(t-3)y(t-1), x(t-3)y(t-2), x(t-3)y(t-3) \}^T
 \end{aligned} \quad (36)$$

which has 28 members. The detected time-varying ERR-Causality is illustrated in Fig. 3.(a) which clearly shows the results are consistent with the expected causality. The corresponding strength is shown in Fig. 3.(b), which illustrates the consistently strong strength during interactions. To study the flexibility of the new ERR-Causality test in term selection, the candidate term set was deliberately chosen to be insufficient

$$\{1, x(t-1), x(t-2), x(t-3), y(t-1), y(t-2), y(t-3)\}^T \quad (37)$$

where all nonlinear terms have been removed and now only 7 linear terms were considered. The detected values of the ERR-Causality test, illustrated in Fig. 3.(c), are relatively accurate even though a significant nonlinear term with a large contribution was not considered, and a complete parameter set was not estimated. The corresponding strength of the causality is shown in Fig. 3.(d), which illustrates the strength during interactions is not as strong as shown in Fig. 3.(b) due to the absence of the non-linear term, but is still distinctive enough to reflect the original causality. A comparison between Fig. 2.(b) and Fig. 3.(c) shows the robustness of the ERR-Causality test compared to the Granger test. It is well understood that the estimation accuracy for both methods can be improved with an increasing number of trials, but for a real system, multi-trials is often impossible.

Note, the start and end positions of  $Y$  causes  $X$  and  $X$  causes  $Y$  in Fig. 3 are not exactly the same as the original model, which is not surprising because the window size determines the reaction speed of causality detection. A selection of small window size means a fast reaction to the change of causality over time, but may lead to insufficient data to achieve an accurate result. Conversely a selection of a large window size can improve the accuracy of causality detection, but may significantly slow down the reaction to the change of causality over time.

### C. Example 3

This example aims to explore the application of the proposed method in the estimation of the time shift between two signals, and compares the performance to the cross correlation method using the same window size. Assuming  $X$  causes  $Y$  at time  $t$ , the time shift is approximated by the time lag of the first term from past information of  $X$  appearing in the detected significant terms ranked by ERR. For example, if  $x(t-3)$  is the first selected term, the time shift of  $X$  causing  $Y$  at time  $t$  is 3 times the sample interval. The contribution of the first term can also be used to approximate the strength of the causality at that time shift.

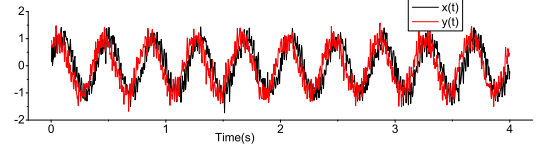


Fig. 4. The generated signal  $x(t)$  and  $y(t)$  for Example 3.

Consider two signals

$$\begin{aligned} x(t) &= \sin(2\pi f_1 t) + 0.2\sin(2\pi f_2 t) + \zeta_x(t) \\ y(t) &= \sin(2\pi f_1 t + \beta) + 0.2\sin(2\pi f_2 t) + \zeta_y(t) \end{aligned} \quad (38)$$

Because the next example involves real EEG data, two common frequencies which appear in real EEG experiments were introduced in this simulation. One is the dominant frequency from EEG, which is typically around  $2-3\text{Hz}$ , and will be denoted by  $f_1$ ; another is  $50\text{Hz}$  induced from electrical interference, denoted by  $f_2$ . Obviously,  $y(t)$  has a fixed phase shift  $\beta$  in front of  $x(t)$  at all time. From the causality point of view,  $Y$  causes  $X$  at all time with a fixed time shift, which can be expressed  $\tau = \frac{\beta}{2\pi \times f_1}$ . Model (38) was simulated by setting the parameters  $f_1 = 2.5\text{Hz}$ ,  $f_2 = 50\text{Hz}$ ,  $\Delta t = 0.004\text{sec}$ ,  $\beta = 0.2\pi$  and  $\zeta_x(t)$  and  $\zeta_y(t)$  are white noise. It can be calculated that the original time shift  $\tau$  equals  $40\text{ms}$ . Fig. 4 shows the signals  $X$  and  $Y$  with noise standard deviation  $\sigma(\zeta_x(t)) = 0.2$  and  $\sigma(\zeta_y(t)) = 0.2$ . The candidate terms set for the ERR-Causality test was chosen as

$$\{1, x(t-\Delta t), \dots, x(t-15\Delta t), y(t-\Delta t), \dots, y(t-15\Delta t)\}^T \quad (39)$$

The detected time shifts from the ERR-Causality test and the cross correlation method using different window sizes are shown in Fig. 5, where it is expected that the accuracy of estimated  $\tau$  for both methods will improve with an increase in window size. Comparison of the results with the same window size for the two methods clearly indicates that the ERR-Causality test requires less samples of data to achieve the same accuracy for  $\tau$ . This means that the ERR-Causality test has faster reactions to the change of time shift over time than that of the cross correlation method. Several tests suggested that the selection of a window size of  $0.5-0.8$  times the period of the dominant wave always produces the best performance.

## IV. APPLICATION TO EEG DATA

Recently, more and more studies have investigated the problem of causal effects in neural data [21, 22], the investigation of which is usually carried out by

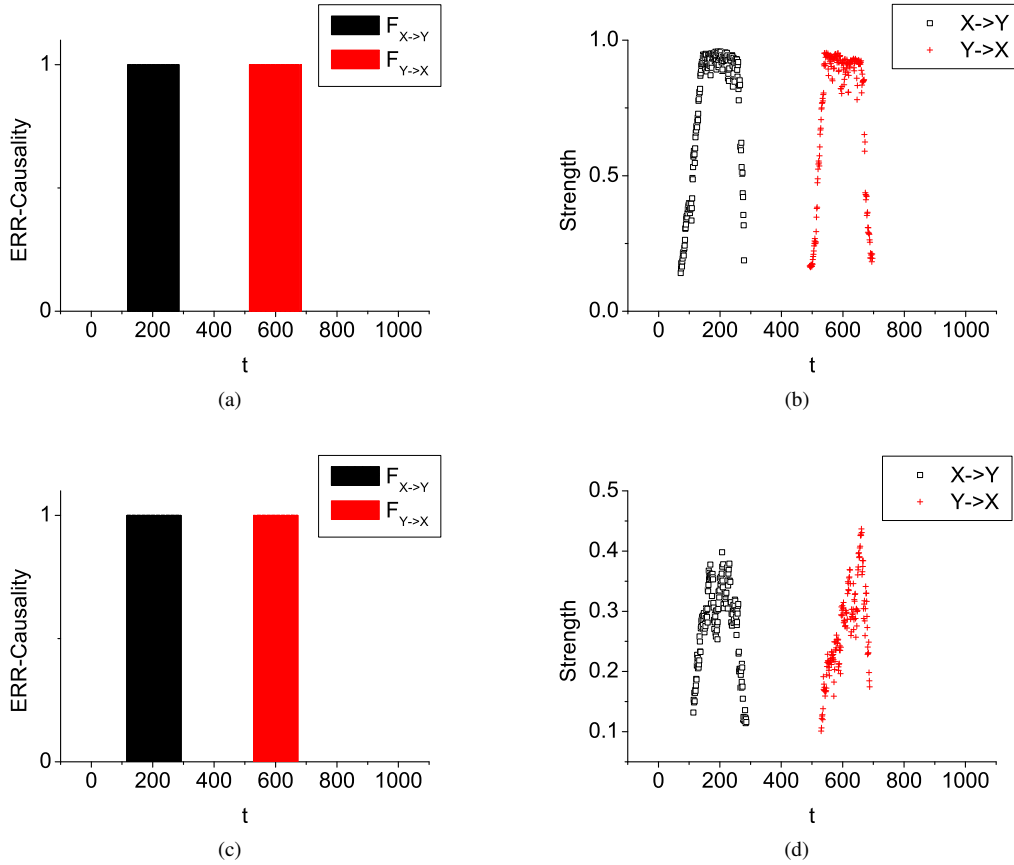


Fig. 3. (a)-(b) The detected time-varying causality with corresponding strength based on the candidate terms set (36) using the ERR-Causality test for Example 2. (c)-(d) The detected time-varying causality with corresponding strength based on the candidate terms set (37) using the ERR-Causality test for Example 2.

correlation and coherence measures [23, 24, 25] or phase synchronisation measures [26, 27]. These methods measure the strength of the interactions between signals, but no insight into the directionality of information flow is produced. The Granger method has also been used to understand the directed interactions between neural assemblies [6], but again no quantitative description has been presented to measure the information flow. In this example, a data set consisting of an epileptic sample of scalp EEGs recorded by the EEG Laboratory of Neurophysiology, Sheffield Teaching Hospitals NHS Foundation Trust, Royal Hallamshire Hospital, were studied to find out the directional flow of signals collected from different cortical sites, and to determine the corresponding quantitative measurements of time shift to try to better understand the functional organization of the brain during an epileptic seizure.

In this example, to simplify the problem, only dominant causality is considered at a specific time by comparing

the strength of both causalities.

#### A. Data acquisition

Scalp EEG signals are synchronous discharges from cerebral neurons detected by electrodes attached to the scalp. A NeuroScan Medical System (NeuroSoft Inc., Sterling, VA) with the international 10-20 electrode coupling system was used. The sampling rate of the device was 500 Hz. A total of 32 EEG series were recorded in parallel from 32 electrodes located on an epileptic seizure patient's scalp using the same 32 channel amplifier system using bipole montage reference channels. This example considered four bipolar montages:  $F7-F3$ ,  $T5-O1$ ,  $F8-F4$ ,  $T6-O2$ , which are located in different sites of the brain, as illustrated by Fig. 6. The montage  $F7-F3$  represents the voltage difference between the channel  $F7$  and  $F3$ . The purpose of this example is to detect the causality associated with the corresponding time shift between the signals from the front and back site

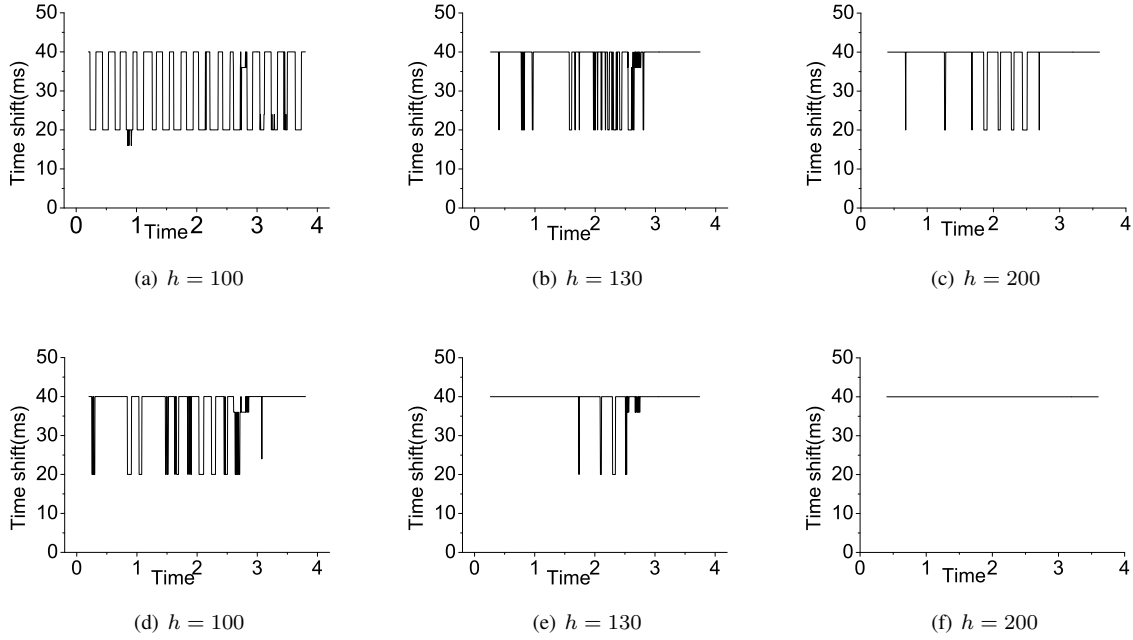


Fig. 5. (a)-(c): The detected time shift of  $Y$  in front of  $X$  using the cross correlation method with different sizes of window  $h$ ; (d)-(f): The detected time shift of  $Y$  in front of  $X$  using the ERR-Causality test with different sizes of window  $h$ .

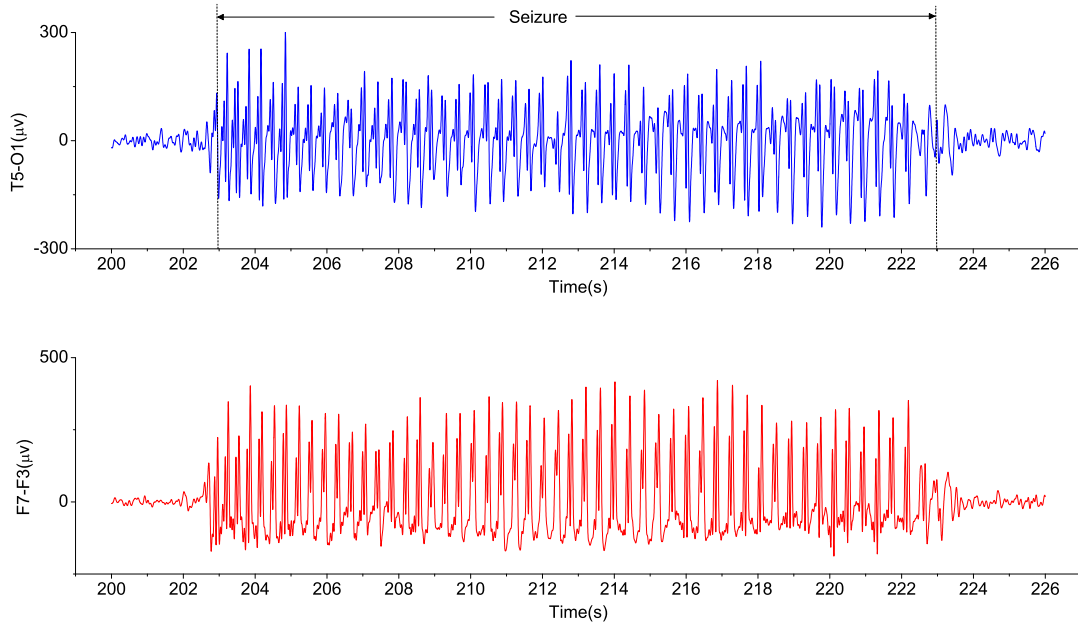


Fig. 7. The recorded EEG signals from the left brain.

of the brain. A comprehensive seizure of a patient was sampled (13000 data points) starts from the 200<sup>th</sup> sec to the 226<sup>th</sup> sec. Two experiments were implemented,

in the first the montages  $F7-F3$  and  $T5-O1$ , two signals from the left brain, were sampled after noise removal and the data are shown in Fig. 7. Fig. 7 clearly shows the

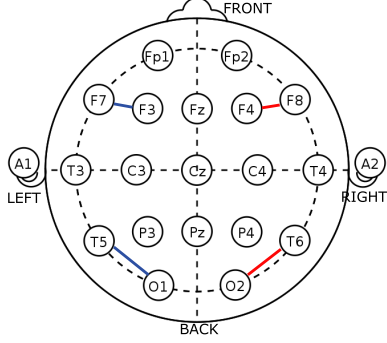


Fig. 6. Distribution of EEG channels in the brain.

epileptic seizure where regular oscillation starts at the  $203^{rd}$  sec and ends at the  $223^{rd}$  sec. Apply the ERR-Causality test, the candidate term set was chosen as

$$\{1, x(t - \Delta t), \dots, x(t - 20\Delta t), y(t - \Delta t), \dots, y(t - 20\Delta t)\}^T \quad (40)$$

and  $\Delta t = 2ms$ . The window size was chosen as 300, which will depend on the dominant frequency of the signals as suggested in Example 3. Fig. 8.(a) shows the contribution of the first term from the past information of the other signal detected by the proposed approach, where the black scattering denotes the strength of the signal  $F7-F3$  causing  $T5-O1$ , and the red scattering denotes the strength of the signal  $T5-O1$  causing  $F7-F3$ . The corresponding values of ERR-Causality test between these two signals are shown in Fig. 8.(b). Inspection of both figures shows that during the time interval  $200.5 - 202$  sec, before the epileptic seizure,  $F7-F3$  causing  $T5-O1$  dominates the interactions. During the time interval  $202 - 223$  sec,  $T5-O1$  causing  $F7-F3$  dominates the interaction, although  $F7-F3$  causing  $T5-O1$  appears occasionally with very short duration, especially during  $202 - 212$  sec the strength of  $T5-O1$  causing  $F7-F3$  is relatively higher and the causality is more consistent than that during  $212 - 223$  sec. During the time interval  $223 - 226$  sec, after the seizure, two causalities appear alternatively with relatively small strength. The detected time shift of the signal  $T5-O1$  in front of  $F7-F3$  is shown in Fig. 8.(c). It is observed that, during the stable procedure of the epileptic seizure (time intervals  $203 - 223$  sec), the detected time shift of the signal  $T5-O1$  in front of  $F7-F3$  is very consistent (the average value is about  $28ms$ ), although a few gaps appear indicating when the opposite causality is detected. From the causality point of view, this observation indicates the signal  $T5-O1$  may cause  $F7-F3$  during the seizure with an averaged time shift of about  $28ms$ .

In the second experiment, the montages  $F8-F4$  and  $T6-O2$ , two signals from the right brain, were sampled after noise removal and the data are shown in Fig. 9. Using the same settings of the parameters, the results produced by the new approach are illustrated by Fig. 10. Fig. 10.(a) shows the contribution of the first term from the past information of the other signal detected by the proposed approach, where the black scattering denotes the strength of the signal  $F8-F4$  causing  $T6-O2$ , and the red scattering denotes the strength of the signal  $T6-O2$  causing  $F8-F4$ . The corresponding values of ERR-Causality test between these two signals are shown in Fig. 10.(b). Inspection of both figures shows that during the time interval  $200 - 202$  sec, before the epileptic seizure, two causalities appear alternatively with relatively small strength. During the time interval  $202 - 223$  sec,  $T6-O2$  causing  $F8-F4$  completely dominates the interaction with relatively higher strength. During the time interval  $223 - 226$  sec, after the seizure, two causalities appear alternatively again with relatively small strength. The detected time shift of the signal  $T6-O2$  in front of  $F8-F4$  is shown in Fig. 10.(c). The observations are very similar as those of the first experiment. During the stable interval of the seizure, the detected time shift of the signal  $T6-O2$  in front of  $F8-F4$  is relatively consistent. Before the start and after the end of the seizure the time shift appears to be chaotic or random. This observation indicates the signal  $T6-O2$  may cause  $F8-F4$  during the seizure with an averaged time shift of about  $23ms$ .

Four more epileptic seizures from the same patient have also been studied and the observations of causality are very similar, and the averaged time shifts during the seizure are very close, as shown in TABLE IV, which demonstrates that the time shift of the considered two signals is within the range of  $25 - 32ms$ .

TABLE IV  
THE DETECTED AVERAGED TIME SHIFT FOR 5 SEIZURES FROM THE SAME PATIENT.

Interval	$\tau(T5-O1 \rightarrow F7-F3)$	$\tau(T6-O2 \rightarrow F8-F4)$
14 - 40s	27.46ms	25.19ms
202 - 223s	28.03ms	22.90ms
560 - 583s	30.04ms	27.90ms
1361 - 1386s	31.31ms	30.32ms
1674 - 1694s	31.16ms	29.55ms

All above results produced by the ERR-Causality test indicate the signals from the back brain dominantly causes the signals from the front brain during an epileptic seizure for the studied patient. Moreover, the time shifts of the signal in the left back brain which is in front of the signal in the left front brain were observed to be very

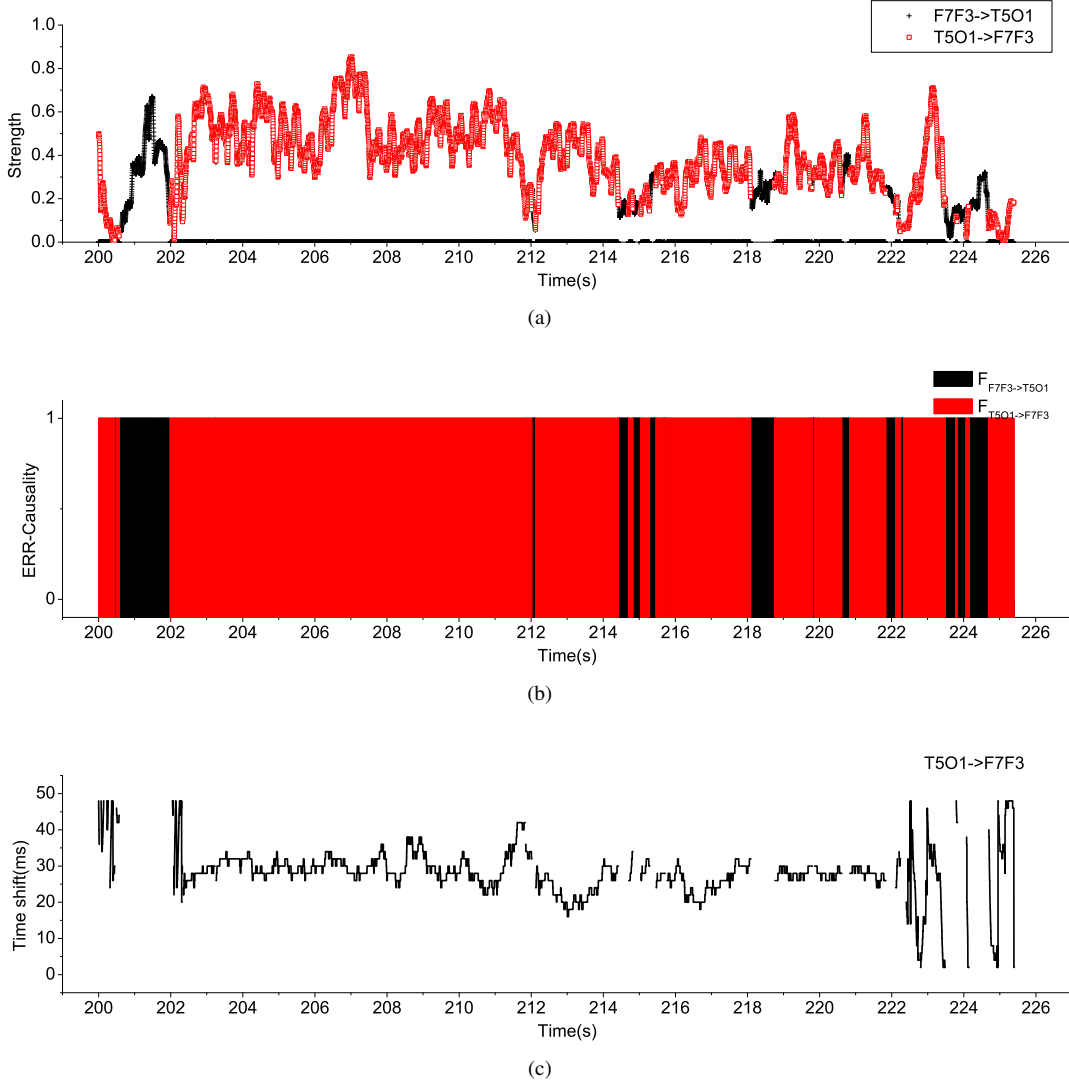


Fig. 8. The results produced by the presented approach for the signal  $F7-F3$  and  $T5-O1$ . (a) The strength of the ERR-Causality, where the black scattering represents  $F7-F3$  causing  $T5-O1$  and the red scattering represents  $T5-O1$  causing  $F7-F3$ ; (b) The detected map of the time-varying causality, where black denotes  $F7-F3$  causing  $T5-O1$  and red denotes  $T5-O1$  causing  $F7-F3$ ; (c) The detected time-varying time shift of the signal  $T5-O1$  in front of  $F7-F3$ .

close to the time shifts of the signal in the right back brain in front of the signal in the right front brain. For all five epileptic seizures studied in this example,  $\tau(F7-F3 \rightarrow T5-O1)$  are slightly different, but consistently longer than  $\tau(F8-F4 \rightarrow T6-O2)$ .

The observations extracted from EEG data are very interesting and may provide significant potential in future studies of brain activity during an epileptic seizure.

## V. CONCLUSIONS

We have shown that the new ERR-Causality test can detect the time-varying causality of two signals,

a measure of the corresponding coupling strength, and estimate the time shift. Both the ERR-Causality test and the Granger test can detect causality accurately if the candidate terms are chosen appropriately. However, the ERR-Causality test has a better performance if some model terms are omitted and estimates of the full parameter set in the model is not required. It has also been demonstrated in Example 3 that the new method has a faster reaction to the change of causality over time than the cross correlation method. The performance of the Granger test was not compared because it is not able to detect the time shift. The application of the ERR-

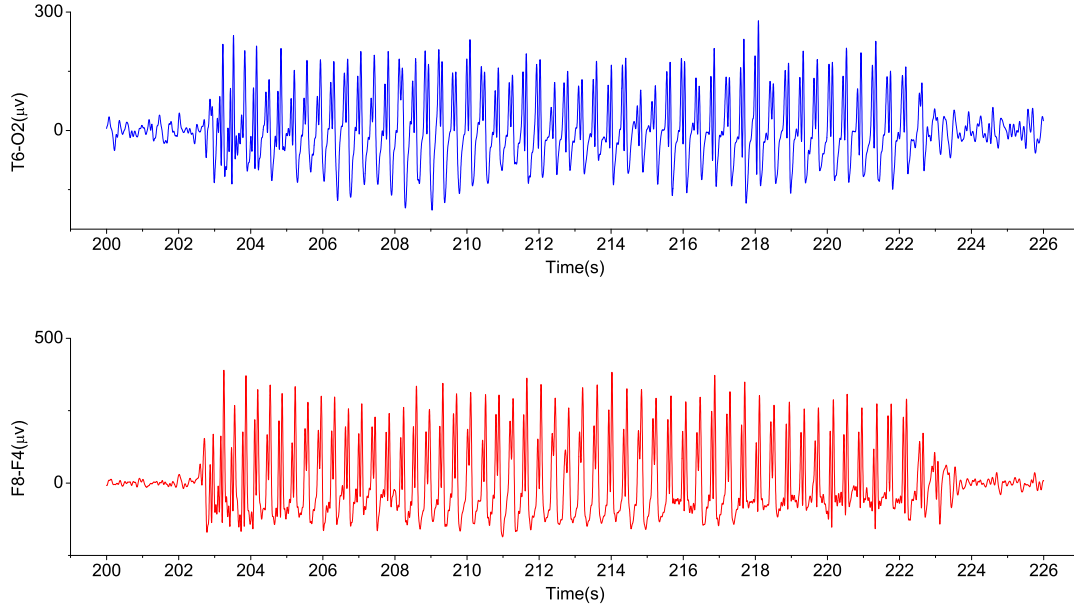


Fig. 9. The recorded EEG signals from the right brain.

Causality test to detect the directed interaction between EEG signals has been presented in the last example. By analysing the detected causality map along with the strength and the estimated time shifts, it has been found that the dominant causality is very consistent during epileptic seizure, but the dominant causality before and at the end of seizure are random. Furthermore, the estimated time shifts of the signals from the back brain causing the signals from the front brain are in the range of  $25 - 32ms$  for the studied patient. These observations show that the proposed method could be a very important tool to help understand the functional organization of the brain during an epileptic seizure by providing an insight into directionality of information flow. The results showing the causality between signals from the back brain and the front brain are highly encouraging, and a full map of signal flow will be developed in future publications.

#### ACKNOWLEDGMENT

The authors gratefully acknowledge that part of this work was financed by Engineering and Physical Sciences Research Council (EPSRC), UK, and by the European Research Council (ERC).

#### REFERENCES

- [1] O. David, S. J. Kiebel, L. M. Harrison, J. Mattout, J. M. Kilner, and K. J. Friston, "Dynamic causal modeling of evoked responses in eeg and meg," *NeuroImage*, vol. 30, pp. 1255–1272, 2006.
- [2] L. A. Baccala, K. Sameshima, G. Ballester, A. C. D. Valle, and C. Timon-laria, "Studying the interaction between brain structures via directed coherence and granger causality," *Applied Signal Processing*, vol. 5, pp. 40–48, 1998.
- [3] L. A. Baccala and K. Sameshima, "Partial directed coherence: a new concept in neural structure determination," *Biological Cybernetics*, vol. 84, pp. 463–474, 2001.
- [4] H. Ombao and S. V. Belleghem, "Evolutionary coherence of nonstationary signals," *IEEE Transactions on Signal Processing*, vol. 55, no. 6, pp. 2259–2266, 2008.
- [5] C. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, pp. 424–438, 1969.
- [6] W. Hesse, E. Moller, M. Arnold, and B. Schack, "The use of time-variant eeg granger causality for inspecting directed interdependencies of neural assemblies," *Journal of Neuroscience Methods*, vol. 124, pp. 27–44, 2003.
- [7] D. Marinazzo, M. Pellicoro, and S. Stramaglia, "Nonlinear parametric model for granger causality of time series," *Physical Review E*, vol. 73, p. 066216, 2006.
- [8] J. P. Hamilton, G. Chen, M. E. Thomason, M. E.



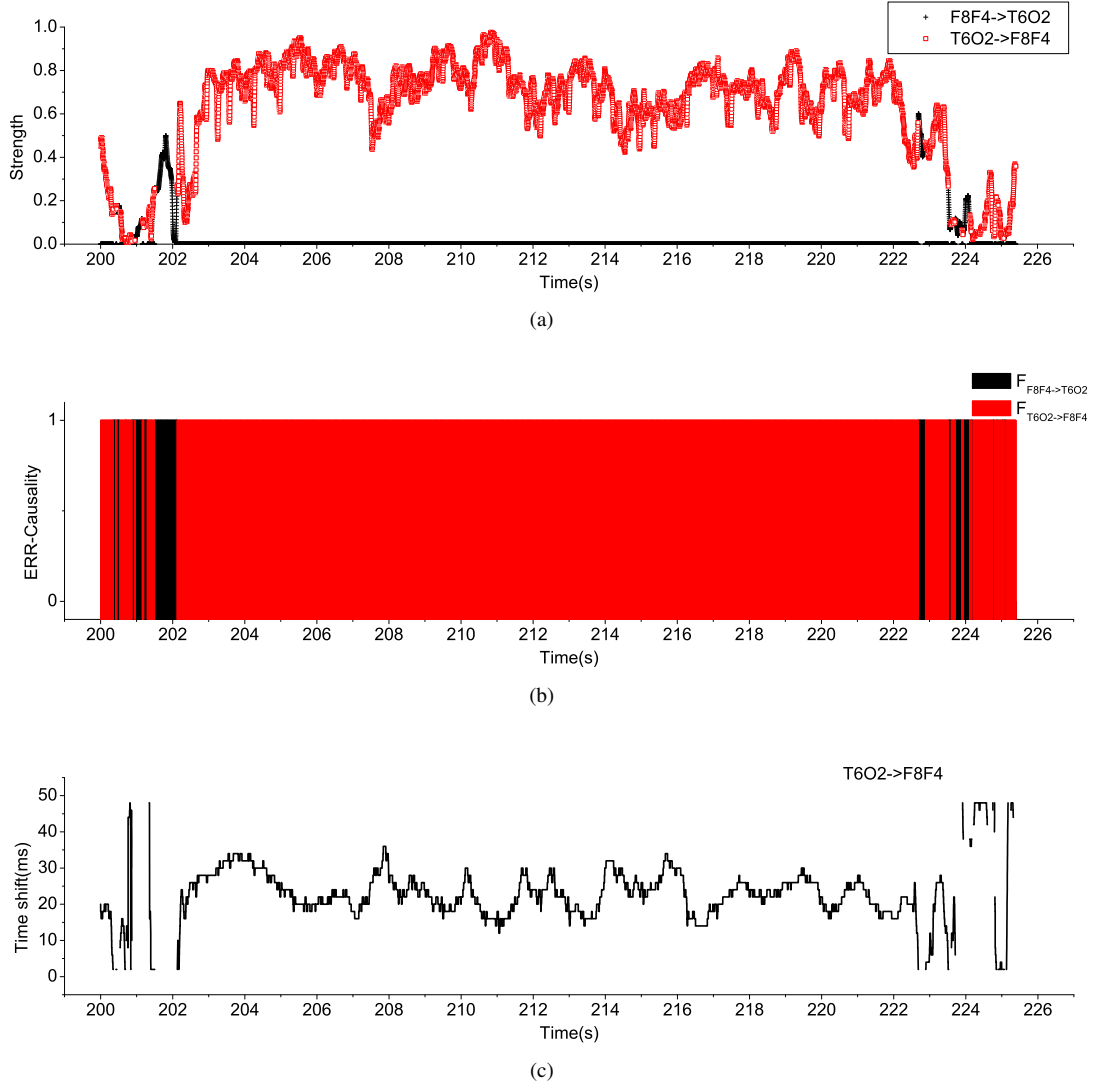


Fig. 10. The results produced by the presented approach for the signal  $F8-F4$  and  $T6-O2$ . (a) The strength of the ERR-Causality, where the black scattering represents  $F8-F4$  causing  $T6-O2$  and the red scattering represents  $T6-O2$  causing  $F8-F4$ ; (b) The detected map of the time-varying causality, where black denotes  $F8-F4$  causing  $T6-O2$  and red denotes  $T6-O2$  causing  $F8-F4$ ; (c) The detected time-varying time shift of the signal  $T6-O2$  in front of  $F8-F4$ .

- Schwartz, and I. H. Gotlib, "Investigating neural primacy in major depressive disorder: multivariate granger causality analysis of resting-state fmri time-series data," *Molecular Psychiatry*, vol. 16, no. 7, pp. 763–772, 2011.
- [9] J. C. Moshier and R. M. Leahy, "Source localization using recursively applied and projected (rap) music," *IEEE Transactions on Signal Processing*, vol. 47, no. 2, pp. 332–340, 1999.
- [10] S. A. Billings, M. Korenberg, and S. Chen, "Identification of non-linear output-affine systems using an orthogonal least squares algorithm," *International Journal of Systems Science*, vol. 19, pp. 1559–1568, 1988.
- [11] M. Korenberg, S. A. Billings, and Y. P. Liu, "Orthogonal parameter estimation algorithm for non-linear stochastic systems," *International Journal of Control*, vol. 48, no. 1, pp. 193–210, 1988.
- [12] S. Chen, S. A. Billings, and W. Luo, "Orthogonal least squares methods and their application to non-linear system identification," *International Journal of Control*, vol. 50, no. 5, pp. 1873–1896, 1989.
- [13] Y. Zhao and S. A. Billings, "Neighbourhood detection using mutual information for the identification



- of cellular automata,” *IEEE Transactions on systems Man and Cybernetics Part B*, vol. 36, no. 2, pp. 473–479, 2006.
- [14] S. A. Billings and H. L. Wei, “Sparse model identification using a forward orthogonal regression algorithm aided by mutual information,” *IEEE Transactions on Neural Networks*, vol. 18, no. 1, pp. 306–310, 2007.
- [15] H. L. Wei, S. A. Billings, Y. Zhao, and L. Z. Guo, “Lattice dynamical wavelet neural networks implemented using particle swarm optimization for spatio-temporal system identification,” *IEEE Transactions on Neural Networks*, vol. 20, no. 1, pp. 181–185, 2009.
- [16] S. A. Billings and H. L. Wei, “An adaptive orthogonal search algorithm for model subset selection and nonlinear system identification,” *International Journal of Control*, vol. 81, no. 5, pp. 714–724, 2008.
- [17] Q. M. Zhu and S. A. Billings, “Fast orthogonal identification of nonlinear stochastic models and radial basis function neural networks,” *International Journal of Control*, vol. 64, no. 5, pp. 871–886, 1996.
- [18] K. Z. Mao and S. A. Billings, “Algorithms for minimal model structure detection in nonlinear dynamic system identification,” *International Journal of Control*, vol. 68, no. 2, pp. 311–330, 1997.
- [19] S. A. Billings, S. Chen, and M. Korenberg, “Identification of mimo non-linear systems using a forward-regression orthogonal estimator,” *International Journal of Control*, vol. 49, no. 6, pp. 2157–2189, 1989.
- [20] Y. Li, H. L. Wei, S. A. Billings, and P. G. Sarigiannis, “Time-varying model identification for time-frequency feature extraction from eeg data,” *Journal of neuroscience methods*, vol. 196, pp. 151–158, 2011.
- [21] A. Brovelli, M. Z. Ding, A. Ledberg, Y. H. Chen, R. Nakamura, and S. L. Bressler, “Beta oscillations in a large-scale sensorimotor cortical network: Directional influences revealed by granger causality,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, pp. 9849–9854, 2004.
- [22] D. W. Gow, J. A. Segawa, S. P. Ahlfors, and F. H. Lin, “Lexical influences on speech perception: A granger causality analysis of meg and eeg source estimates,” *NeuroImage*, vol. 43, pp. 614–623, 2008.
- [23] S. L. Bressler and J. A. S. Kelso, “Cortical coordination dynamics and cognition,” *Trends in Cognitive Sciences*, vol. 5, pp. 26–36, 2001.
- [24] G. Marrelec, J. Daunizeau, M. Pelegrini-Issac, J. Doyon, and H. Benali, “Conditional correlation as a measure of mediated interactivity in fmri and meg/eeg,” *IEEE Transactions on Signal Processing*, vol. 53, no. 9, pp. 3503–3516, 2005.
- [25] L. Astolfi, F. Cincotti, D. Mattia, F. D. Fallani, G. Vecchiato, S. Salinari, G. Vecchiato, H. Witte, and F. Babiloni, “Time-varying cortical connectivity estimation from noninvasive, high-resolution eeg recordings,” *Journal Of Psychophysiology*, vol. 24, no. 2, pp. 93–90, 2010.
- [26] F. Varela, J. P. Lachaux, E. Rodriguez, and J. Martinerie, “The brainweb: phase synchronization and large-scale ingeration,” *Nature Reviews Neuroscience*, vol. 2, no. 4, pp. 229–239, 2001.
- [27] S. Aviyente and A. Y. Mutlu, “A time-frequency-based approach to phase and phase synchrony estimation,” *IEEE Transactions on Signal Processing*, vol. 59, no. 7, pp. 3086–3098, 2011.